

BAB II

TINJAUAN PUSTAKA DAN DASAR TEORI

2.1. Tinjauan Pustaka

Diasrina Dahri (2016) melakukan pembuatan aplikasi perangkat lunak sistem pendukung keputusan untuk menentukan kelayakan calon penerima Bantuan Bidikmisi di Universitas Mulawarman dengan menerapkan metode *Naïve Bayes*. Bantuan Bidikmisi yang diterapkan dalam penelitian ini adalah Bantuan Bidikmisi yang diperuntukkan bagi para siswa SMA/SMK/MA/MAK berprestasi namun kurang mampu yang telah lulus jalur SNMPTN, SMMPTN. Kriteria yang digunakan yaitu Nama Mahasiswa, Pekerjaan Orang Tua, Penghasilan Orang Tua, Jumlah Tanggungan, Daya Listrik (Watt), Nilai Ujian Nasional, dan Keterangan. Perbedaan penelitian ini dengan penelitian yang diusulkan yaitu pada kriteria dan jenis Bantuan Bidikmisi yang digunakan. Bantuan Bidikmisi Pengganti diperuntukkan bagi mahasiswa kurang mampu dan berprestasi yang belum memperoleh beasiswa dengan menggantikan posisi penerima Bantuan Bidikmisi sebelumnya yang mencabut Bantuan Bidikmisi dengan alasan yang tertera pada Pedoman Bidikmisi 2017, sedangkan Bantuan Bidikmisi diperuntukkan bagi para siswa SMA/SMK/MA/MAK berprestasi namun kurang mampu yang telah lulus jalur SNMPTN, SMMPTN, dan SBMPTN. Penelitian ini menghasilkan sebuah aplikasi sistem pendukung keputusan dengan tingkat akurasi sebesar 85.56%.

Antony Anwari Rahman dan Agus Suryanto (2017) melakukan penelitian implementasi sistem informasi seleksi penerima beasiswa dengan metode *Naïve Bayes Classifier*. Perangkat dibuat dengan *framework Microsoft Visual Studio*

2013. Uji kelayakan dilakukan dengan pengujian 1 dan 2. Hasil penelitian dari pengujian 1 sebesar 96,56%, dan pengujian 2 sebesar 90,33%, termasuk dalam kategori sangat layak. Pada penelitian ditemukan bahwa sistem menggunakan algoritma *Naïve Bayes Classifier* menghasilkan akurasi yang tinggi.

Mulyadi (2016) melakukan penelitian penerapan algoritma *Naïve Bayes* untuk klasifikasi penerima beasiswa. Data yang digunakan untuk klasifikasi merupakan data primer yaitu dataset beasiswa pada Universitas BSI. Dari hasil pengujian model yang dilakukan, diperoleh nilai akurasi yang tinggi untuk klasifikasi penerima beasiswa dengan nilai akurasi 100% dan AUC 1.

Fitri Atun (2017) melakukan penelitian melalui skripsi berjudul Implementasi Algoritma *Naïve Bayes Classifier* untuk Memprediksi Penjurusan di SMA Negeri 3 Bantul. Program diujikan pada data pelatihan (*training*) dengan jumlah data 152 record. Kemudian uji coba prediksi penjurusan dengan data pengujian (*testing*) dengan jumlah data sebesar 38 *record*, diperoleh rata-rata tingkat persentase ketelitian sebesar 78,0701%.

Tabel 2. 1 Perbandingan Hasil Penelitian

Parameter Peneliti	Objek	Metode	Hasil
Diasrina Dahri, (2016)	Mahasiswa pendaftar Bantuan Bidikmisi	Algoritma <i>Naïve Bayes</i>	Menghasilkan sebuah aplikasi sistem pendukung keputusan dengan tingkat akurasi sebesar 85,56%.
Antony Anwari Rahman dan Agus Suryanto (2017)	Data penerima beasiswa di suatu lembaga pendidikan	Algoritma <i>Naïve Bayes</i> untuk meningkatkan akurasi dan efisiensi dan	Uji kelayakan dilakukan dengan pengujian 1 dan 2. Hasil penelitian dari pengujian 1

		metode <i>Research and Development</i> (R&D) untuk metode penelitian	sebesar 96,56% dan pengujian 2 sebesar 90,33%.
Mulyadi (2016)	Data penerima beasiswa prestasi Universitas BSI Bandung tahun 2015	Algoritma <i>Naïve Bayes</i> dan metode ROC untuk mencari AUC	Hasil pengujian model yang dilakukan, diperoleh nilai akurasi tinggi untuk klasifikasi penerima beasiswa yaitu <i>Naïve Bayes</i> dengan nilai akurasi 100% dan AUC 1.
Fitri Atun (2017)	Siswa kelas X SMA Negeri 3 Bantul	Algoritma <i>Naïve Bayes</i>	Uji coba prediksi penjurusan dengan data pengujian (<i>testing</i>) dengan jumlah data sebesar 38 <i>record</i> , diperoleh rata-rata tingkat persentase ketelitian sebesar 78,0701%.
Siti Mutiah (Usulan)	Mahasiswa pendaftar Bantuan Bidikmisi pengganti	Algoritma <i>Naïve Bayes</i>	Algoritma <i>Naïve Bayes</i> diharapkan mampu memberikan prediksi kelayakan calon penerima beasiswa bidikmisi pengganti di Universitas Gadjah Mada.

2.2. Dasar Teori

Dasar teori digunakan untuk memahami definisi, pengertian dasar dan istilah yang digunakan dalam penelitian ini. Berikut dasar teori yang digunakan:

2.2.1. Bantuan Bidikmisi Pengganti

Bidikmisi merupakan salah satu program unggulan pemerintah di bawah naungan Kementerian Riset Teknologi dan Pendidikan Tinggi (Kemenristekdikti) yang pelaksanaannya sudah dimulai sejak tahun 2010. Jumlah peminat Program Bidikmisi menunjukkan peningkatan yang sangat signifikan dari tahun ke tahun, untuk tahun 2017 tercatat sebanyak 520.688 pelamar tetapi hanya sekitar 80.000 saja yang bisa diakomodir karena keterbatasan anggaran pemerintah.

Dijelaskan di dalam pedoman Program Bidikmisi Tahun 2018 bahwa penerima bantuan Bidikmisi yang tidak dapat melanjutkan program dengan alasan:

1. Dikeluarkan sebagai mahasiswa dari Perguruan Tinggi;
2. Mengundurkan diri;
3. Meninggal dunia; atau
4. Diberhentikan sebagai penerima Bidikmisi oleh Pengelola karena melanggar ketentuan Bidikmisi Perguruan Tinggi

maka bantuan Bidikmisi dapat dialihkan kepada mahasiswa lain yang angkatan, jenjang pendidikannya sama, dan sesuai dengan kriteria penerima Bantuan Bidikmisi.

Adapun persyaratan bagi para calon pendaftar Bantuan Bidikmisi Pengganti (Pedoman Bidikmisi, 2018):

1. Tidak mampu secara ekonomi dengan kriteria:
 - a. Penerima Beasiswa Siswa Miskin (BSM) atau Pemegang Kartu Indonesia Pintar (KIP) atau sejenisnya; atau
 - b. Pendapatan kotor orang tua/wali gabungan (suami + istri) setinggi-tingginya Rp4.000.000,00 (Empat juta rupiah) atau pendapatan kotor gabungan orang tua/wali dibagi jumlah anggota keluarga maksimal Rp750.000,00 (Tujuh ratus lima puluh ribu rupiah).
2. Penerima Bidikmisi Pengganti ditetapkan oleh Perguruan Tinggi setiap tahun akademik.
3. Mahasiswa aktif dan sedang menjalani perkuliahan pada semester normal.
4. Memiliki potensi akademik baik dibuktikan dengan transkrip nilai akademik.

2.2.2. Data Mining

Data mining merupakan proses penggunaan teknik statistik, kecerdasan buatan atau *artificial intelligence* (AI), pengenalan pola, dan sistem basis data yang digunakan untuk memanfaatkan data dalam basis data yang diolah sehingga menghasilkan informasi baru yang berguna. Berikut beberapa pengertian data mining yang secara naratif mempunyai beberapa maksud yang mirip (Prasetyo, 2014):

- a. Pencarian otomatis pola dalam basis data besar, menggunakan teknik komputasional campuran dari statistik, pembelajaran mesin, dan pengenalan pola.
- b. Pengekstrakan implisit non-trivial, yang sebelumnya belum diketahui secara potensial adalah informasi berguna dari data.
- c. Ilmu pengekstrakan informasi yang berguna dari set data atau basis data besar.
- d. Proses penemuan informasi otomatis dengan mengidentifikasi pola dan hubungan tersembunyi dalam data.

Pekerjaan yang berkaitan dengan data mining dapat dibagi menjadi empat kelompok:

- a. Model Prediksi (*Prediction Modelling*)

Pekerjaan ini berkaitan dengan pembuatan sebuah model yang dapat melakukan pemetaan dari setiap himunan variabel ke setiap targetnya, kemudian menggunakan model tersebut untuk memberikan nilai target pada himpunan baru yang didapat. Ada 2 jenis model prediksi, yaitu klasifikasi dan regresi. Klasifikasi digunakan untuk variabel target diskret, sedangkan regresi digunakan untuk variabel target kontinu.

- b. Analisis *Cluster* (*Cluster Analysis*)

Analisis *Cluster* melakukan pengelompokan data ke dalam sejumlah kelompok berdasarkan kesamaan karakteristik masing-masing data pada kelompok-kelompok yang ada. Analisis Asosiasi (*Association Analysis*)

Analisis Asosiasi digunakan untuk menemukan pola yang menggambarkan kekuatan hubungan fitur dalam data. Pola yang ditemukan merepresentasikan bentuk aturan implikasi atau subset fitur.

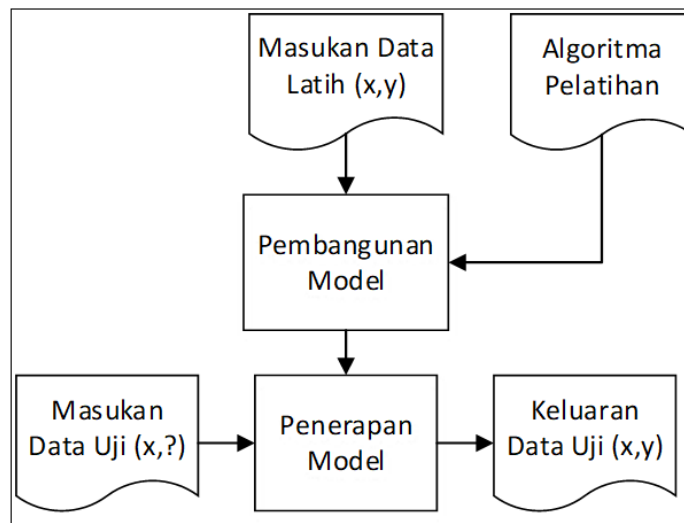
c. Deteksi Anomali (*Anomaly Detection*)

Pekerjaan deteksi anomaly berkaitan dengan pengamatan sebuah data dari sejumlah data yang secara signifikan mempunyai karakteristik yang berbeda dari sisa data yang lain.

2.2.3. Klasifikasi

Klasifikasi merupakan proses pembelajaran suatu fungsi tujuan (target) f yang memetakan tiap himpunan atribut x ke satu dari label kelas y yang didefinisikan sebelumnya. Fungsi target disebut juga model klasifikasi. Teknik klasifikasi cocok untuk memprediksi atau menggambarkan data-set dengan tipe biner atau nominal. Teknik ini kurang efektif digunakan untuk tipe data ordinal karena teknik ini tidak mempertimbangkan secara implisit urutan dalam kategori data (Hermawati, 2013).

Kerangka kerja klasifikasi ditunjukkan pada Gambar 2.1. Pada gambar tersebut, disediakan sejumlah data latih (x,y) untuk digunakan sebagai data pembangun model, kemudian menggunakan model tersebut untuk memprediksi kelas dari data uji $(x,?)$ sehingga data uji $(x,?)$ diketahui kelas y yang seharusnya (Prasetyo, 2014).



Gambar 2. 1 Proses pekerjaan klasifikasi

2.2.4. Algoritma *Naïve Bayes*

Algoritma *Naive Bayes* merupakan salah satu algoritma yang terdapat pada teknik klasifikasi. *Naive Bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris *Thomas Bayes*, yaitu memprediksi peluang di masa depan berdasarkan pengalaman dimasa sebelumnya sehingga dikenal sebagai *Teorema Bayes*. Teorema tersebut dikombinasikan dengan *Naive* dimana diasumsikan kondisi antara atribut saling bebas. Klasifikasi *Naive Bayes* diasumsikan bahwa ada atau tidak ciri tertentu dari sebuah kelas tidak ada hubungannya dengan ciri dari kelas lainnya (Bustami, 2014).

Persamaan dari teorema *Bayes* adalah:

$$P(C|X) = \frac{P(X|C).P(C)}{P(X)} \dots\dots\dots(2.1)$$

Keterangan:

X : Data dengan *class* yang belum diketahui

C : Hipotesis data X merupakan suatu *class* spesifik

$P(C/X)$: Probabilitas hipotesis C berdasarkan kondisi X (*posteriori probability*)

$P(C)$: Probabilitas hipotesis C (*prior probability*)

$P(X/C)$: Probabilitas X berdasarkan kondisi pada hipotesis C

$P(X)$: Probabilitas X

Untuk menjelaskan teorema *Naive Bayes*, perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas apa yang cocok bagi sampel yang dianalisis tersebut. Karena itu, teorema *Bayes* pada persamaan (2.1) disesuaikan menjadi persamaan (2.2) berikut:

$$P(C|F_1 \dots F_n) = \frac{P(C)P(F_1 \dots F_n|C)}{P(F_1 \dots F_n)} \dots\dots\dots (2.2)$$

Dimana Variabel C merepresentasikan kelas, sementara variabel $F_1 \dots F_n$ merepresentasikan karakteristik petunjuk yang dibutuhkan untuk melakukan klasifikasi. Maka rumus tersebut menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam kelas C (*Posterior*) adalah peluang munculnya kelas C (sebelum masuknya sampel tersebut, seringkali disebut *prior*), dikali dengan peluang kemunculan karakteristik – karakteristik sampel pada kelas C (disebut juga *likelihood*), dibagi dengan peluang kemunculan karakteristik – karakteristik sampel secara global (disebut juga *evidence*). Karena itu, rumus diatas dapat pula ditulis secara sederhana pada persamaan (2.3):

$$Posterior = \frac{Prior \times likelihood}{evidence} \dots\dots\dots (2.3)$$

Nilai *Evidence* selalu tetap untuk setiap kelas pada satu sampel. Nilai dari *posterior* tersebut nantinya akan dibandingkan dengan nilai – nilai *posterior* kelas lainnya untuk menentukan ke kelas apa suatu sampel akan diklasifikasikan. Penjabaran lebih lanjut rumus *Bayes* tersebut dilakukan dengan menjabarkan $(C|F_1, \dots, F_n)$ menggunakan aturan perkalian sebagai berikut:

$$\begin{aligned} P(C|F_1, \dots, F_n) &= P(C) P(F_1, \dots, F_n|C) \\ &= P(C)P(F_1|C)P(F_2, \dots, F_n|C, F_1) \end{aligned}$$

$$\begin{aligned}
&= P(C) P(F_1|C) P(F_2|C, F_1) P(F_3, \dots, F_n|C, F_1 F_2) \\
&= P(C) P(F_1|C) P(F_2|C, F_1) P(F_3|C, F_1, F_2), P(F_4, \dots, F_n|C, F_1, F_2, F_3) \\
&= P(C) P(F_1|C) P(F_2|C, F_1) P(F_3|C, F_1, F_2) \dots P(F_n|C, F_1, F_2, F_3, \dots, F_{n-1}) \dots (2.4)
\end{aligned}$$

Dapat dilihat bahwa hasil penjabaran tersebut menyebabkan semakin banyak dan semakin kompleksnya faktor – faktor syarat yang mempengaruhi nilai probabilitas, yang hampir mustahil untuk dianalisa satu persatu. Akibatnya, perhitungan tersebut menjadi sulit untuk dilakukan. Disinilah digunakan asumsi independensi yang sangat tinggi (naif), bahwa masing – masing petunjuk (F_1, F_2, \dots, F_n) saling bebas (independen) satu sama lain. Dengan asumsi tersebut, maka berlaku suatu persamaan sebagai berikut:

$$P(F_i|F_j) = \frac{P(F_i \cap F_j)}{P(F_j)} = \frac{P(F_i)P(F_j)}{P(F_j)} = P(F_i) \dots\dots\dots (2.5)$$

Untuk $i \neq j$, sehingga

$$P(F_i|C, F_j) = P(F_i|C) \dots\dots\dots (2.6)$$

Dari persamaan di atas dapat disimpulkan bahwa asumsi independensi naif tersebut membuat syarat peluang menjadi sederhana, sehingga perhitungan menjadi mungkin untuk dilakukan. Selanjutnya, penjabaran $P(C|F_1, \dots, F_n)$ dapat disederhanakan menjadi:

$$\begin{aligned}
P(C|F_1, \dots, F_n) &= P(C)P(F_1|C)P(F_2|C)P(F_3|C) \dots P(F_n|C) \\
&= P(C) \prod_{i=1}^n P(F_i|C) \dots\dots\dots (2.7)
\end{aligned}$$

Keterangan:

$\prod_{i=1}^n P(F_i|C)$ = Perkalian rating antar atribut

Persamaan di atas merupakan model dari teorema *Naïve Bayes* yang selanjutnya akan digunakan dalam proses klasifikasi. Untuk klasifikasi dengan data kontinyu digunakan rumus Densitas Gauss:

$$P(X_i = x_i | Y = y_j) = \frac{1}{\sqrt{2\pi}\sigma_{ij}} e^{-\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2}} \dots\dots\dots (2.8)$$

Keterangan

P = Peluang

X_i = Atribut ke-i

x_i = Nilai Atribut ke-i

Y = Kelas yang dicari

y_i = Sub-kelas yang dicari

μ = mean, menyatakan rata-rata dari seluruh atribut

σ = Deviasi Standar, menyatakan varian dari seluruh atribut

Naive Bayes merupakan metode yang menggunakan pendekatan probabilitas untuk menghasilkan klasifikasi. Metode ini menggabungkan probabilitas *term* dengan probabilitas kategori untuk menentukan kemungkinan kategori berhasil. Dari penjelasan mengenai algoritma *Naive Bayes*, dapat disimpulkan langkah – langkah pengerjaan *Naive Bayes* adalah (Diasrina, 2016):

Tentukan data latih dan data uji yang ingin diklasifikasikan. Menghitung $P(C_i)$ yang merupakan probabilitas *prior* untuk setiap sub kelas C yang dihasilkan menggunakan persamaan:

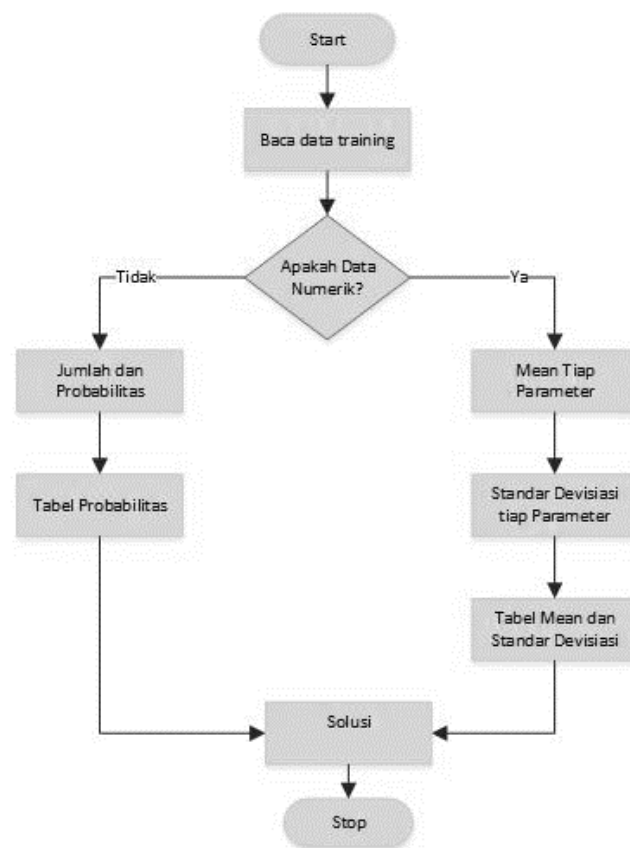
$$P(C_i) = \frac{S_i}{s} \dots\dots\dots (2.9)$$

Dimana S_i adalah jumlah data *training* dari kategori C_i , dan s adalah jumlah total data *training*. Menghitung $P(F_i|C_i)$ yang merupakan probabilitas *posterior* F_i dengan syarat C menggunakan persamaan (2.7). Apabila F_i merupakan data numerik, maka untuk menghitung $P(F_i|C_i)$ menggunakan distribusi Gaussian yang terdapat pada persamaan (2.8).

Memaksimalkan $P(F_i|C_i)P(C_i)$ untuk mendapatkan kelas C yang ingin diklasifikasikan dengan cara mengkalikan $P(F|C_i)$ dan $P(C_i)$ untuk semua kemungkinan klasifikasi:

$$P(C_i) \prod_{i=1}^n P(F_i|C_i) \dots\dots\dots (2.10)$$

Dengan kata lain, hasil yang ditetapkan ke dalam kelas C_i adalah yang mempunyai $P(F_i|C_i)P(C_i)$ maksimum.



Gambar 2. 2 Skema *Naive Bayes*

Alur dari metode *Naive Bayes* dapat dilihat pada Gambar 2.2 di atas

(Bustami, 2014). Adapun keterangan dari Gambar 2.2 sebagai berikut:

1. Baca data *training*
2. Hitung jumlah dan probabilitas, namun apabila data numerik maka:

- a. Cari nilai *mean* dan standar deviasi dari masing-masing parameter yang merupakan data numerik. Adapun persamaan yang digunakan untuk menghitung nilai rata – rata (*mean*) dapat dilihat sebagai berikut:

$$\mu = \frac{\sum_{i=1}^n x_i}{n}$$

atau

$$\mu = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} \dots\dots\dots (2.11)$$

di mana:

μ = rata – rata hitung (mean)

x_i = nilai sampel ke - i

n = jumlah sampel

Adapun persamaan untuk menghitung standar deviasi adalah sebagai berikut:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n-1}} \dots\dots\dots (2.12)$$

di mana:

σ = standar deviasi

x_i = nilai x ke - i

μ = rata – rata hitung

n = jumlah sampel

- b. Jika data berbentuk kategorikal cari nilai probabilitas dengan cara menghitung jumlah data yang sesuai dari kategori yang sama dibagi dengan jumlah data pada kategori tersebut. Adapun persamaan untuk menghitung nilai probabilitas bagi data kategorikal adalah sebagai berikut:

$$P(X_i = x_i | Y = y_i) = \frac{\sum \text{nilai data pada kelas } X_i}{\sum \text{data kelas } X_i} \dots\dots\dots (2.13)$$

di mana:

P = peluang

X_i = nilai atribut

x_i = atribut ke $-i$

Y = kelas yang dicari

y_i = Sub kelas Y yang dicari

3. Mendapatkan nilai dalam tabel *mean*, standard deviasi dan probabilitas

Penghitungan probabilitas data numerik menggunakan densitas Gauss pada persamaan 2.8

4. Solusi kemudian dihasilkan

5. Menghitung akurasi sistem

Untuk mengevaluasi performa dari model yang dibangun, perlu dilakukan pengukuran performa, yaitu pengukuran akurasi (*accuracy*) atau tingkat kesalahan (*error rate*). Jika f_{ij} menotasikan jumlah *record* dari kelas I yang berada di kelas j pada saat pengujian, maka pengukuran akurasi (*accuracy*) dapat ditulis dengan persamaan sebagai berikut (Hermawati, 2013):

$$Accuracy = \frac{\text{jumlah_prediksi_yg_benar}}{\text{jumlah_prediksi_keseluruhan}} \times 100\% = \frac{f_{11}+f_{00}}{f_{11}+f_{10}+f_{01}+f_{00}} \dots\dots\dots (2.14)$$

Sedangkan tingkat kesalahan (*error rate*) didefinisikan sebagai berikut:

$$Error = \frac{\text{jumlah_prediksi_yg_salah}}{\text{jumlah_prediksi_keseluruhan}} \times 100\% = \frac{f_{01}+f_{10}}{f_{11}+f_{10}+f_{01}+f_{00}} \dots\dots\dots (2.15)$$

2.2.5. WEB

World Wide Web atau *www* atau juga dikenal dengan WEB adalah salah satu layanan yang didapat oleh pemakai komputer yang terhubung ke internet. Secara makna sebuah website adalah sekumpulan halaman informasi yang disediakan melalui jalur internet sehingga bisa diakses diseluruh dunia selama terkoneksi dengan jaringan internet tanpa terbatas ruang dan waktu. *Web* adalah

suatu metode untuk menampilkan informasi di internet, baik teks, gambar, suara maupun video yang interaktif dan mempunyai kelebihan untuk menghubungkan (*link*) satu dokumen dengan dokumen yang lainnya (*hypertext*) yang dapat diakses melalui sebuah *browser*. Tahapan-tahapan pengembangan *website* dari awal hingga akhir, adalah *planning*, *design*, *developing/coding*, *upload web*, *promote*, *maintenance web*. Fungsi *website* secara umum adalah fungsi komunikasi, fungsi informasi, fungsi intertainment, dan fungsi transaksi.

2.2.6. MySQL

MySQL adalah salah satu aplikasi pengolahan *database* yang sering digunakan oleh banyak *programmer*, disamping sudah mendukung dalam pembuatan *database* yang berbasis *Client/Server*, juga dapat mengolah *database* dalam jumlah besar. SQL adalah bahasa stansar untuk *query* yang digunakan untuk manipulasi data yang dibuat dalam berbagai DBMS (*Database Management System*), salah satunya adalah MySQL. Beberapa keistimewaan yang dimiliki oleh MySQL adalah *Portability*, *Open Source*, *Multiuser*, *Performent Tuning*, *Column Tuning*, *Command dan Function*, *Security*, *Scolabitily dan Limits*, *Connectivity*, *Localization*, *Interface*, *Cliens dan Tools*, *Structure Table*.